CRU TS 3.10 - Data availability and File Formats explained

Where to find the CRU TS 3.10 data and metadata files:

The CRU TS 3.10 data (i.e. climate variables) and metadata files (i.e. stations and observations) are available from the BADC Archive at:

http://badc.nerc.ac.uk/browse/badc/cru/data/cru_ts/cru_ts_3.10/data

where the 'data' directory contains the CRU TS 3.10 monthly gridded variables.

In July 2012, a systematic error was discovered in the CRUTS v3.10 process. The effect was, in some cases, to reduce the gridded values. The files (for pre, frs and wet) were immediately removed from BADC and from the Climate Explorer at KNMI. The corrected run, based on the v3.10 precipitation station data, should be used as a direct replacement. The version number 3.10.01 reflects this status. No replacement files were provided for wet and frs variables as CRU TS 3.2 was about to be released).

All the data files (ASCII ".dat" and netcdf ".nc") are compressed (.gz extension).
Note: The files are >2GB when unzipped, which may cause problems on the computer systems of some users.

CRU TS 3.10 Data and Metadata File Formats explained:

The CRU TS data are stored in both ASCII and NetCDF formats:

- **ASCII data**: The 360-lat x 720-long grid is presented exactly as that, with 720 columns, and 360 rows per timestep. The first row in each grid is the southernmost (centred on 89.75S). The first column is the westernmost (centred on 179.75W). There are scaling factors in use in the data files (see Table below). One gets the whole global grid for the first time step, then the whole grid for the second, and so on. So the first 360 rows show the data for Jan 1901, next 360 rows the data for Feb 1901, next 360 rows for March 1901 and so on.

- **NetCDF data**: CRU3.10 - All attributes are correctly set (unlike CRU 3.00). There are no scale factors in the NetCDF files because the data is FLOAT instead of INT.

  Please see the CEDA NetCDF pages at http://www.ceda.ac.uk/help/users-guide/file-formats/netcdf/ for more information.

- Missing values are stored as -999.

How to read the CRU TS 3.10 data:

The CRU TS 3.10 data files contain:

Label	**Variable**	**Units**(Multiplying factor for ASCII data ONLY)	**Comments**
cld	Cloud Cover	percentage (x10)	Cloud is derived from measurements of Sun Hours. Cloud cover data is also synthesised from DTR (Diurnal Temperature Range) in areas where sun hours are not measured.
dtr	Diurnal Temperature Range	Degrees Celcius (x10)	
pet	Potential Evapo-Transpiration (PET)	Millimetres (x10)	The method used is the FAO (Food and Agricultural Organization) grass reference evapotranspiration equation (Ekstrom et al., 2007, which is based on Allen et al., 1994). It is a variant of the Penman Monteith method using the gridded TMP, TMN, TMX, VAP and CLD. Note that PET values are mean mm/day for each month (with a scaling factor of 10 applied to the PET ascii (*.dat) files, but NOT the PET netcdf files (*.nc). The pet values in the datafiles therefore need to be muliplied by the number of days for each month to get the mean pet for that month
pre	Precipitation	Millimetres (x10)	See CRU TS 3.10.01 precipitation data files
tmp	Daily mean temperature	Degrees Celcius (x10)	
tmn	Monthly average daily minimum temperature	Degrees Celcius (x10)	
tmx	Monthly average daily maximum temperature	Degrees Celcius (x10)	
vap	Vapour pressure	Hecta-Pascals (x10)	

To read the CRU TS 3.10 ASCII data, users have so far been writing their own scripts as these are fairly easy to parse. If you would like to share your script to read the CRU TS ASCII data with other users, then please email BADC Support. The ASCII data should be read using free-format.

To read the CRU TS3.10 NetCDF data, you may use any NetCDF enabled software, such as Xconv, CDAT or FERRET).

How to read the CRU TS 3.10 Station data (metadata):

The CRU TS 3.10 station files are available from the BADC Archive at:http://badc.nerc.ac.uk/browse/badc/cru/data/cru_ts/cru_ts_3.10/station/

There are now two kinds of station files. Both types contain one value for every value in the data file:

- **Regular '.stn.' files**. The values in these represent, for each cell and timestep, the number of station that could have influenced the data value for that cell and timestep. The sphere of influence is the Correlation Decay Distance (CDD), which is 450 km for precipitation, 750 km for diurnal temperature range, and 1200 km for mean temperature (New et al, 2000).

- **Cell station '.sn0.' files**. These new files give the actual number of all stations reporting in that cell at that timestep.

Station data files are available for the following variables: pre, cld ('stn' files only), dtr, tmpdtr (for dtr, tmn and tmx), tmp and vap. There are no station data files available for pet, wet and frs.

Please note that due to an encoding error, the station count files ('sn0' and 'stn') will need to be **multiplied by a factor of 10** to obtain their true integer values.

There is also an elevation file available in the data directory (halfdesg.elv.grid.data.gz)

All the station data files (ASCII ".dat" and netcdf ".nc") are compressed (.gz extension).

How to read the CRU TS 3.21 Observation data (metadata):

The CRU TS datasets are built from databases of observations of primary variables. The Daily Mean Temperature (TMP), Monthly average daily minimum temperature (TMN), Monthly

average daily maximum temperature (TMX) and Precipitation (PRE) observations are made available, corresponding to the 3.10 (3.10.01 for PRE) release of CRU TS:

http://badc.nerc.ac.uk/browse/badc/cru/data/cru_ts/cru_ts_3.10/observation/

All the data files are ASCII ".dtb"

The format of the observations is as follows:

Each file contains a set of station records, one after the other. A typical record consists of a header line, a normals line, and then a line of data for each year of observations.

The header line consists of the following fields:

WMO code	i7	country (2), station (3), optional (2)
Latitude	i5	degrees x 100
Longitude	i6	degrees x 100
Altitude	i4	metres
Station	a20	name of station
Country	a13	name of country
Start	i4	earliest year of observations
End	i4	latest year of observations

Each field is separated from the next by a single space.

The normals line is no longer used by the CRU TS processes and may be ignored.

Data lines start with the year (i4), followed by twelve monthly observations (12i5).

For TMP, values are in degrees C x 10.

For PRE, the values are in mm x 10.

Missing values are indicated thus: -9999.

The first four lines of the first record in the v3.10 PRE database are:

```
-511900  6100   1060  190 BIRI            NORWAY      1895 1992
6190  449  344  378  373  564  712  866  909  860  888  704  485
1895-9999-9999-9999-9999-9999-9999 1410 1850  670  800  860  400
1896  140   40  850  140  190 1050 1080  810  890 1450  230  350
```

The WMO code is -511900, the negative sign indicating a temporary or unknown WMO code. The latitude is 61N and the longitude 10.6E, with an altitude of 190m. The station is BIRI, in NORWAY, and the data run from 1895 to 1992.

The normals line is ignored.

The first data line is for 1895. January to June are missing. The values for the rest of the year are:

```
 July       141mm
 August     185mm
 September   67mm
 October     80mm
 November    86mm
 December    40mm
```

The second data line is for 1896, and values may be calculated for all twelve months (January being 14mm).

Note 1: These database files are functionally equivalent to the ones used in production of the CRU TS dataset. They are, however, smaller. Only stations that were actually used in the gridding process are included; stations with insufficient data between 1961 and 1990 are excluded, as are those with missing location information. This mirrors the checking done at the start of the CRU TS update process, and is indicated by the word 'clean' in the filenames.

Note 2: The term 'database' is being used loosely here, to describe collections of monthly station observation records in flat text files. The format of the files is fixed, to maintain backwards compatibility with previous programs, and dates from an age when data storage was highly limited and techniques such as integer recording of real values were commonplace.